AD-A032 693 MARYLAND UNIV COLLEGE PARK DEPT OF COMPUTER SCIENCE F/G 5/7
THE REPRESENTATION AND SELECTION OF COMMONSENSE KNOWLEDGE FOR N--ETC(U)
MAY 76 C RIEGER N00014-76-C-0477
UNCLASSIFIED TR-458 NL

| OF |

AD
A032693

END
DATE
FILMED
1-77

AD A032693

NR-049-386
code 437

12

COMPUTER SCIENCE
TECHNICAL REPORT SERIES

D D C
RECEIVED
NOV 29 1976
C

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742

14 TR-458

15 N00014-76C-0477 new

11 May 1976

12 31p.

6

The Representation and Selection of Commonsense Knowledge for Natural Language Comprehension

10 Chuck Rieger

Department of Computer Science
University of Maryland
College Park, Md. 20742

Revision of paper presented at Georgetown University Linguistics Roundtable; to appear in Roundtable Proceedings, 1976

The research described in this report was funded by the Office of Naval Research, under contract number N00014-76C-0477.

ABSTRACT Representation and selection of commonsense knowledge about cause and effect are central aspects of human intelligence. This report describes a representation technique called Commonsense Algorithms, and describes the organization of knowledge encoded in this scheme. Current research is focussed on three major areas: the application of the representation formalism to (1) ~~(a)~~ the comprehension of a children's story called The Magic Grinder; (2) the representation and simulation of devices and mechanisms, and (3) problem solving. Selectional issues are discussed with respect both to the commonsense algorithm representation, and to word sense selection in natural language parsing.

I. Introduction

All problems of human intelligence modeling - including natural language analysis - seem eventually to devolve onto problems of how to represent and access general world knowledge. It has been the tradition of most research in natural language to attempt to excise as much of the world as possible in order to focus on such seemingly better defined subproblems of the language phenomenon as "syntax", "semantics" and "pragmatics". Implicit in this approach is the attitude - perhaps "hope" is a better word - that some components of language are indeed meaningful to study in isolation. I will not attempt to argue the soundness of this point of view one way or the other here. Instead, it will be the purpose of this paper to propose how we ought to approach the problems of representing and accessing general world knowledge, under the assumption that such knowledge will be useful to all aspects of language analysis.

Specifically, this paper will focus on two issues which seem to be fundamental to any model of human cognitive abilities: (1) a representation formalism which will address "static" problems of how to express knowledge about the world in the form of memory patterns, and (2) selection processes, as the primary "dynamic" component of intelligence. Rather than address purely language related questions, the discussion will be aimed at a broader spectrum of human abstract symbol processing which will embrace such other so-called intelligent human behavior as "problem solving", "story comprehension" and "learning".

Although the discussion of representation ought logically to precede the selectional issues, it seems appropriate first to motivate the representation by identif ing more precisely what is intended by the term "selection". I want to propose that there are exactly two very abstract, qualitatively distinct classes of selection with which any symbol processing intelligence must ultimately contend:

(1) STRATEGY SELECTION: an ability to select from among a finite collection of alternative courses of action for achieving some desired goal, based

on the context, the goal itself, and a knowledge of what would consitute an "appropriate" solution, or response

(2) THING SELECTION: an ability to select a particular individual for inclusion in some strategy. Things are non-algorithmic concepts like physical objects, other people, word senses, thoughts, etc.

Strategy selection - choosing some algorithmic process to achieve a goal - obviously must be couched upon a very powerful representation of knowledge about cause and effect in the world. If one acknowledges - as we will - that any individual possesses a very large number of relatively small patterns expressing a teleological knowledge of the world across the broad spectrum of human experience - from tying shoe laces to formulating international political strategies - then the act of selection from among such a vast resource of patterns during any act of problem solving or interpretation of event sequences must inevitably play a central role. Hence, in section II a formalism for expressing patterns of cause and effect, and a framework for intelligent selection among them, will be proposed.

While algorithm selection relates to notions of best strategy for accomplishing a task, thing selection is the embodiment of an intelligent decision making component that allows the symbol processor to commit itself to particular objects or concepts during the course of plan synthesis or interpretation. Examples of this kind of selection are somewhat more diverse than algorithmic selection, particularly in the context of a natural language understanding system. Word sense selection during parsing is a very important category of thing selection, where the "things" range over the possible senses of each word in the sentence being parsed in a given context. Parsing is thus viewable as a multiple act of thing selection. Referent identification - the mapping of an object's external description onto a unique internal concept which represents the object in a model - is another thing selection process that manifests itself during the comprehension of text. In the realm of problem solving, during both plan synthesis and plan execution, the system, having committed itself to a particular strategy, will eventually have to commit itself to particular objects in order to carry out the various components of the strategy, e.g. particular hammers, people, insults, and so forth.

If we choose to view intelligence from a perspective that elevates selection to the top of the theory, then it becomes central to any intelligent system that selection be performed in an orderly, and as informed a way as possible at every stage. Experience from the Artificial Intelligence point of view has

been that failure to acknowledge the importance of intelligent selection at every decision point in any computation can lead to run-time-wise cataclysmic and often theoretically vacuous results. Traditional theories of purely grammatical parsing as a class are a case in point.

I want to return to discuss a computational model of intelligent strategy and thing selection. However, let us now turn to the representational formalism which will underlie both forms of selection.

II. CSA Representation

The representation formalism is called the Commonsense Algorithm (CSA) representation. The philosophy evident in its structure reflects a specific point of view within AI, namely that:

(1) a knowledge of cause and effect at all levels of a model is the basis of processes that exhibit intelligent behavior;

(2) such knowledge can be expressed in explicit, decomposable and rearrangeable patterns which can be treated either as data or as process;

(3) there is a syntax to this knowledge that makes it possible to perform orderly and evolutionary transformations on it which amount to forms of "learning".

The domain of the CSA representation - the entities which the representation connects into cause and effect patterns - consists of instances of five generic types of event: (1) actions, (2) tendencies, (3) states, (4) statechanges, and (5) wants:

ACTIONS: *forces which originate from volitional internal commands within "intelligent" organisms, normally intended to contribute toward the accomplishment of some state or statechange (internally or externally)*

TENDENCIES: forces whose origin is not ultimately internal to a volitional, intelligent organism, and which therefore are not the product of a "decision to act". Tendencies provide a way of incorporating in the formalism a knowledge of natural laws (in commonsense terms) which are <u>force</u>

generators, and hence action-like, but which are not goal-directed

STATES: conditions in the world, or internal to the model, which are not action related ... those aspects which remain if we imagine a totally actorless and tendencyless environment

STATECHANGES: changes in conditions in the world, or internal to the model, along dimensions which are continuous

WANTS: internal states of potential actors which motivate them to perform actions intended ito contribute to the attainment of some other desired state.

It is conjectured that these five generic event types are both necessary and sufficient ingredients for capturing all world knowledge of cause and effect. In the fairly diverse range of cause and effect knowledge the CSA group has been considering during the past year, there is no counter evidence to this conjecture. In a sense, however, this is not surprising, since the five categories are quite abstract.

The other, more distinctive, component of the CSA representation is its set of approximately 30 LINKS which are designed to express commonsense cause and effect relationships among instances of these five generic event categories. By constructing graph-like patterns of events and links, CSA gives us the ability to represent in a computer model a fairly detailed knowledge about causality. Some examples will be shown in a moment.

We would hope eventually to be able to conclude that there are exactly N CSA connective links that are necessary and sufficient for expressing all knowledge about cause and effect, regardless of the domain of the knowledge. We would furthermore hope to be able to assert that the links correspond closely to culture-independent cognitive primitives. (This is in contrast with theories such as Conceptual Dependency which assert the existence of a small set of primitive actions, or events [Schank, 1972]. CSA does not seek a set of primitive actions or events, but rather a primitive syntax for knowledge, defined by the links and rules for connecting them.)

It is a key point of CSA theory that the set of connective links be (a) small in number, (b) descriptively adequate for all human knowledge about cause and effect, and (c) universal, perhaps to the extent that they model some level of the genetic endowment of all normal humans. So far, it is clear only that the links in present use satisfy criterion (a)! However, they have survived applications in rather diverse domains

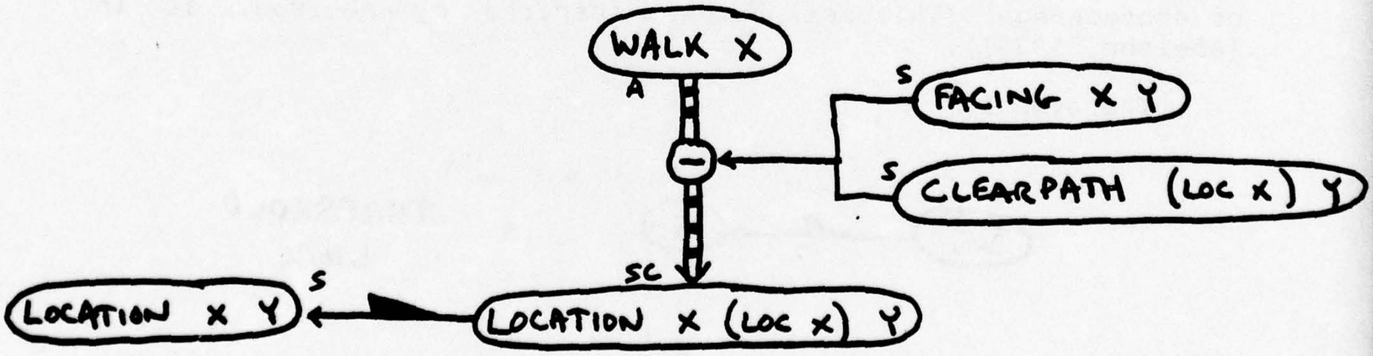
(representing a children's story, expressing the operation of physical devices and mechanisms... a reverse trap flush toilet, an incandescent lightbulb, components of a computer, a computer program, to name a few ... and expressing some simple principles of social and psychological interaction among people). Because of this, we feel that these links constitute some sort of core for any representation.

The CSA representation will be described in this paper only through some examples, with the main goal being to set the stage for the subsequent discussion of selection. [Rieger, 1975 and 1976a] describe the CSA representation in more detail.

The examples about to be shown are intended to be suggestive of the power of a CSA-like approach, and the range of CSA applicability. Bear in mind that, where only one pattern is shown, there will be perhaps thousands of companion or alternative patterns in the computer's memory - each dealing with some part of the world, specific or abstract. The idea will be to organize the patterns in a manner which will cause the higher level memory processes of plan synthesis and language comprehension to see only the most relevant ones at the appropriate times.

<u>EXAMPLE 1</u> -- A very small pattern about locomotion.

Typical of the smallest and most fundamental patterns in CSA are ones such as:



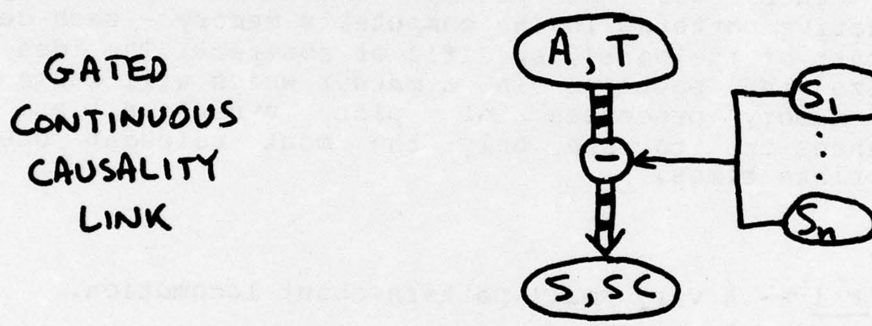
Pattern 1

This CSA pattern illustrates two of the CSA links and has instances of three of the five generic event types. It is read (by a human!) as follows:

"Person X's performance of the "primitive" action (WALK X) will continually cause a statechange in X's location, SC:(LOCATION X (LOC X) Y), from where he is, (LOC X), toward somewhere else, Y, provided that (1) X remains pointed in the right direction, (FACING X Y), and (2) a clear path between X and Y exists for the duration of the activity, namely, (CLEARPATH (LOC X) Y); eventually, such a statechange in location will reach a distinguished level, in this case when X finally reaches Y, (LOCATION X Y)."

In this pattern, (WALK X) is an action, SC:(LOCATION X (LOC X) Y) is a statechange, and (FACING X Y) and (CLEARPATH (LOC X) Y) and (LOCATION X Y) are states.

The symbol



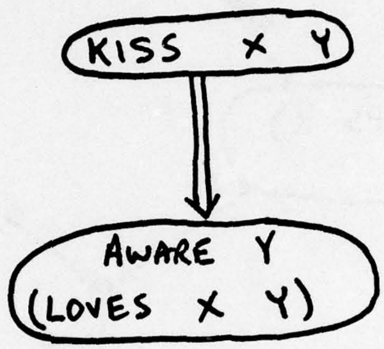
is called the GATED CONTINUOUS CAUSALITY link. This link specifies that action A or tendency T's continued existence will continuously cause state S or statechange SC, provided that other conditions (states) S1,...,Sn are present throughout A's duration. The S1,...,Sn are called GATES, in that they are like valves which control the flow of causality from action to state or statechange. (This metaphor was inspired by Abelson, as in [Abelson, 1973]).

The symbol

is called the THRESHOLD link. This link asserts that statechange SC will eventually reach some distinguished level S; "distinguished" means that the existence of this particular level will affect some other event of interest in the algorithm. In this case, (LOCATION X Y) is of interest because it is the expressed <u>purpose</u> of the algorithm captured by this graph, i.e. to cause X to be located at Y.

In this case, (WALK X) happens to be "primitive" in the sense that it describes an activity which presumably could be implemented directly, in a context-independent way, by robot engineers. However, the CSA graph syntax will actually allow us to include a reference to another entire CSA algorithm wherever an action is needed. This allows us to use more expressive predicates, while retaining the power to elaborate, or define, those predicates in terms of other algorithmic patterns.

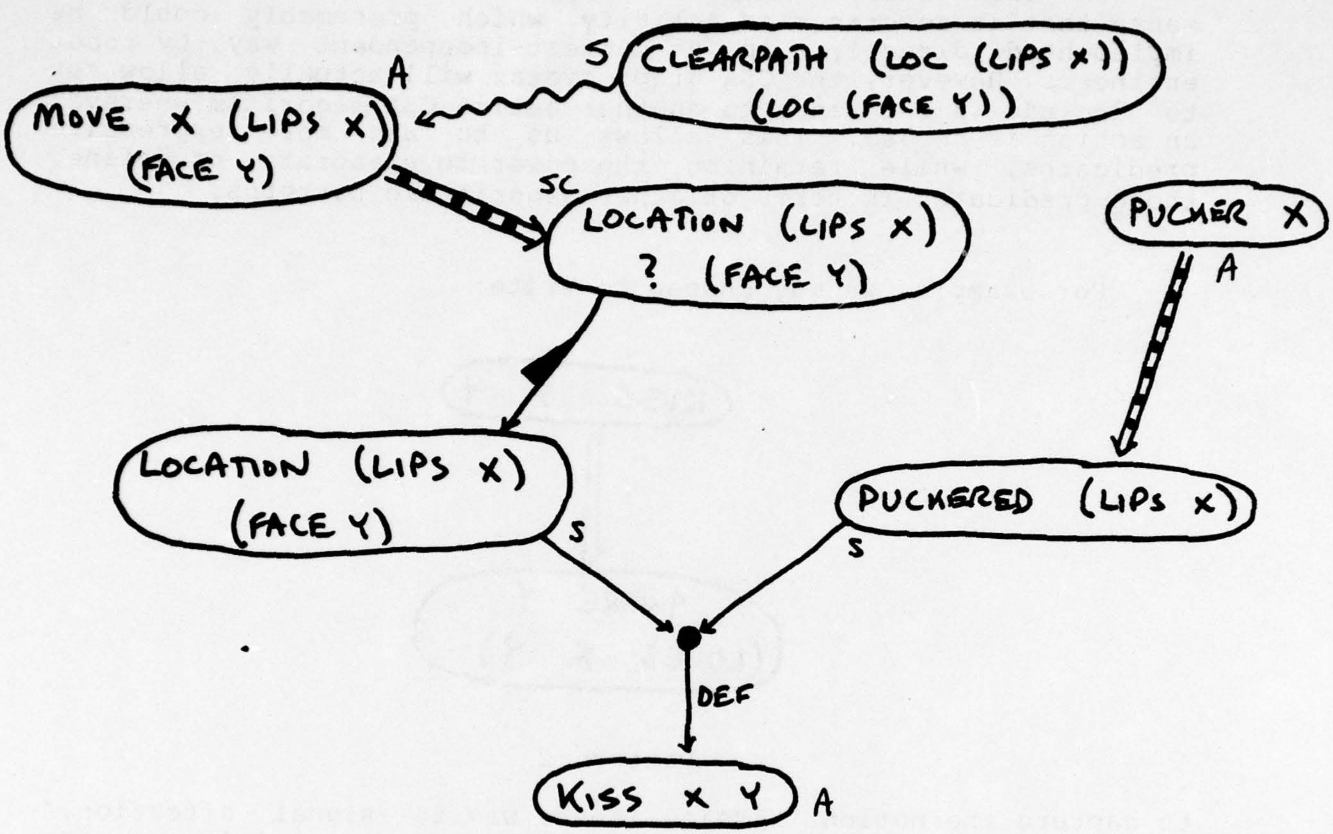
For example, we may choose to write:



Pattern 2

to capture the notion "kissing is one way to signal affection." We may then choose either to regard KISS as a primitive, in the sense that our robot engineer could program his robot directly with a KISS reflex, or to define KISS as simply another algorithmic activity which itself is reducible to other CSA patterns. In this case, we would probably define (KISS X Y) to be simply a compressed way of saying that actor X employed the following strategy:

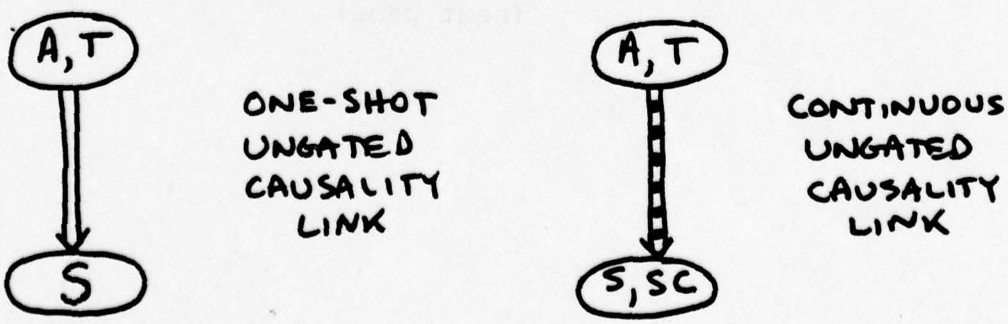
(next page)



Pattern 3

This graph, indexed somewhere else in memory, tells the system that KISS actually refers to a complex goal state consisting of two goals which must be achieved concurrently: X´s lips are puckered, and X´s lips are in physical contact with Y´s cheek. So in case our robot engineer has forgotten to supply us with a KISS primitive reflex, our robot stands a chance of still having a love life!

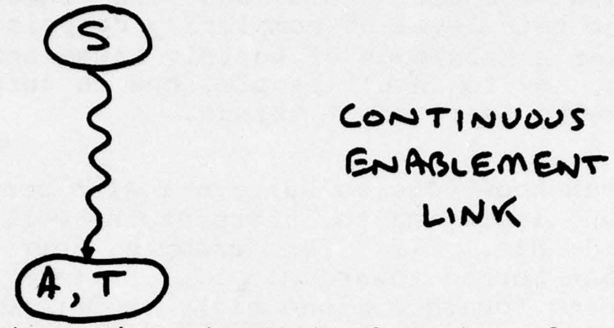
These KISS patterns reveal two other causal links:



called the ONE-SHOT UNGATED CAUSALITY link (action A´s execution

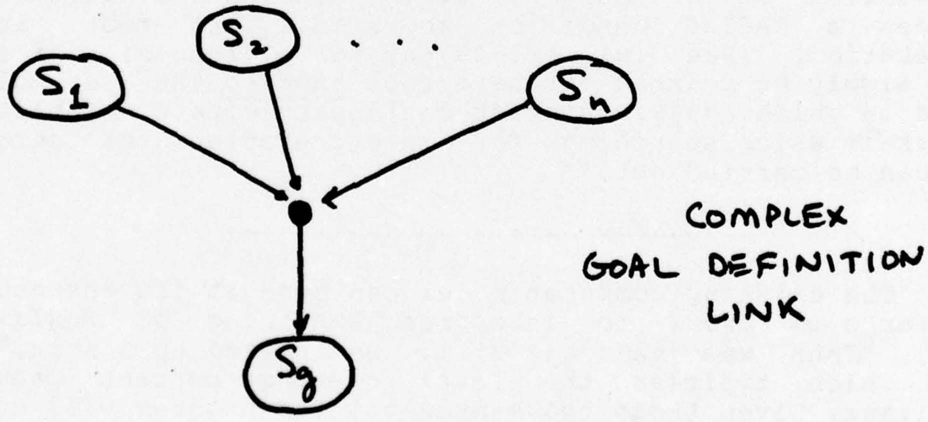
is required only once to achieve state S, and is not required to maintain S), and the CONTINUOUS UNGATED CAUSALITY link which indicates that action A must continually be performed to sustain state S or statechange SC.

Additionally, the CONTINUOUS ENABLEMENT link



has specified that, throughout the duration of action A, state S must remain present to allow A itself to proceed. This notion is distinct from the notion of gates, which govern whether or not an action that is ongoing will achieve the result specified by some causal link. Thus, CSA decouples the prerequisites of performing an action from the prerequisites for the action's achieving some goal.

There is one more link evident in the second KISS pattern. It is called the COMPLEX GOAL DEFINITION link, and is written:



This link couples an arbitrary number of states S1,...,Sn together in order to assert that the goal Sg requires all the specified states to be in effect simultaneously.

Before proceeding, it seems that a case ought to be made for the utility of such patterns in any system that purports to be a model of human intelligence (as manifest, say, in language

comprehension or problem solving). Returning to Pattern 1, we would argue that such a WALK pattern would be quite useful both to a robot that has just been told or decided for itself to go somewhere: namely, that one alternative is first to accomplish a couple of subgoals: (FACING X Y) and (CLEARPATH X Y), then to execute the action (WALK X). This pattern therefore captures one of the possibly numerous strategies for moving from one place to another. In this respect, this and the thousands of other patterns of about this level of complexity comprise the basis of a plan synthesizer's knowledge of worldly cause and effect, e.g., how to go places, how to insult people, how to turn appliances on and use them, how to learn about things.

However, such knowledge as Pattern 1 also bears significance for a robot who is trying to interpret the world and events he perceives around him. If, for example, our robot reads (perceives) "John turned toward Mary..." Pattern 1 would suggest that, since turning toward someone might be intended to achieve a FACING condition, and since a facing condition is part of a WALK pattern, that John just might be getting ready to walk toward Mary, and that John wants to be located at Mary for some reason. Of course, (FACING X Y) may also be a component of a possibly large number of other CSA patterns having nothing to do with going places. Therefore, together with this pattern about walking, the collection of CSA patterns in which the condition (FACING X Y) occurs comprises a set of possible interpretations of FACING events.

CSA theory specifies how such a set of alternate interpretations may be searched to yield a context sensitive interpretation of a thought, i.e. to choose one pattern which references a FACING condition above all the rest as an interpretation. (See [Rieger, 1976a] for a discussion of this.) It will simply be pointed out here that knowing the set of CSA patterns in which any given event could participate establishes a framework in which searching for interpretations of perceived events can be carried out.**

** The existing computer model can consult its inventory of CSA patterns in order to interpret sentences of English in context. "John was mad at Bill. He picked up a rock." is a problem which typifies the level of the current program's capabilities. Given these two sentences, the program will consult patterns about hitting and determine (a) that the referent of "he" in the second sentence is John, and (b) that John was about to propel the rock toward Bill. This is possible because the action of grasping something in the context (a collection of predictions) set up by the first sentence strongly fits into a hitting CSA pattern. [Rieger, 1976a] describes this mechanism in more detail and a forthcoming report will explain the theory at the level of the program which implements it.

The point to be emphasized, therefore, is that a large collection of CSA patterns such as these we have been illustrating will be one very important source of knowledge that underlies both an ability to plan, or solve problems, and an ability to interpret, or draw inferences from a continuous bombardment of perceptions.

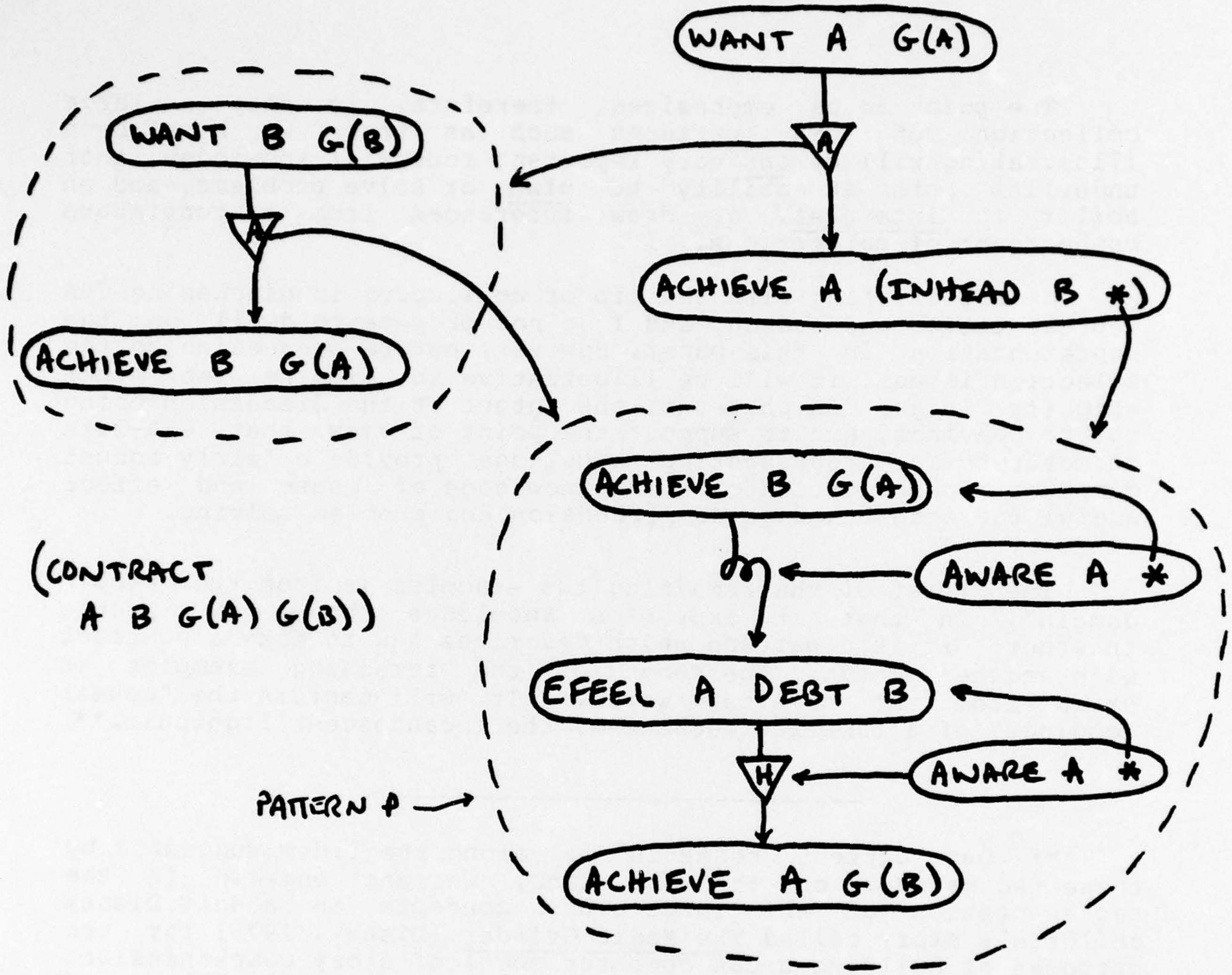
We are flirting with the tip of an iceburg in discussing CSA representation and search, and I do not propose to dwell on the representation in this paper. However, before proceeding to the selection issues, it will be illustrative to examine two other slightly larger CSA patterns, the intent of the discussion being not to convince, but to support the point of view that CSA-like structures for representing knowledge provide a fairly robust computer representation of human knowledge of cause and effect useful for both language comprehension and problem solving.

The first of the remaining two examples is from the "people domain", in that it expresses knowledge about how people interact: it is a pattern which describes how to make a contract with another person. The second of the remaining examples is drawn from the physical world. It will capture the "causal topology" of a familiar mechanism, the incandescent lightbulb.**

** Our current research is along the lines suggested by these two graphs: on the one hand, we are engaged in the representation of the large-scale concepts in a Walt Disney children's story called The Magic Grinder [Disney, 1975] for the purposes of building a CSA computer model of story comprehension. On the other hand, we are investigating a broad spectrum of man-made mechanisms in an attempt to define and delimit human knowledge about cause and effect in mechanical, electrical and computational devices. Although it would seem that such different domains as these two would demand different formulations of cause and effect, we are discovering that such is not the case. Indeed, it is the discovery of common patterns and principles which are domain-independent that constitutes the most exciting aspect of the CSA theory.

EXAMPLE 2: A pattern expressing the concept of a two-person contract, as motivated by The Magic Grinder children's story.

(next page)



Pattern 4

This pattern expresses one general strategy for accomplishing a goal, namely, get someone else to do it by attempting to implant a pattern in his mind which will cause him to behave in the desired way. This single pattern would underly specific instances of contracts ranging from "If you shoot yourself, I'll give you a nice funeral" to "Hughes Aircraft contracted with the government to build 100 jets."

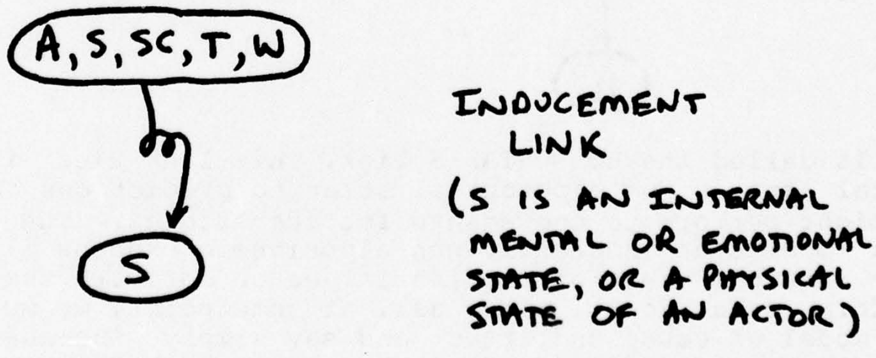
The pattern reads as follows:

> "Person A, wishing to accomplish goal G(A), implants a pattern P in B's mind because A believes that if P were in B's head, B would then do things which would tend to achieve G(A). This pattern to be implanted is: if B achieves G(A), and A is aware of it, such will induce in A a feeling of indebtedness toward B, motivating A to do actions which would tend to accomplish G(B), something A believes B to desire."

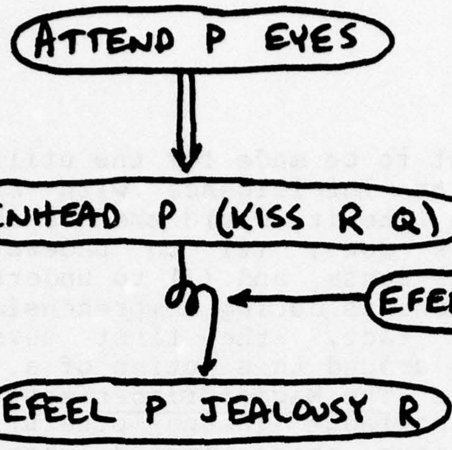
Again, a case ought to be made for the utility of a memory pattern such as this: an intelligence with access to this contract pattern, or one like it, would then be able (1) to make contracts to get its work done, (2) to understand the word "contract" in some deeper sense, and (3) to understand instances of contracts, or its components during comprehension of, say, a children's story. In fact, the first several pages of The Magic Grinder revolve around this notion of a contract. Any model which "understands" The Magic Grinder must surely have to appeal to this type of knowledge (among others) in order to comprehend why the various characters do what they do at the beginning of the story:

> "Once there was a poor maid named Minnie. She worked for the greedy Lord Gurr. While he sat in the shade all day, Minnie and her nephews worked in his garden. Minnie picked fruit and vegetables. Morty and Ferdie pulled and cut weeds. At the end of each day, they brought their basket of food to Lord Gurr. He put the heavy basket on the scale. 'Not bad,' he would say. But whenever Minnie asked for her pay, he always shouted, 'COME BACK TOMORROW!' So Minnie had no money."

The "contract" pattern reveals three more CSA links (in fact, the three most important ones for describing actors and their motivations and intentions). One is called the INDUCEMENT link, and is written as:



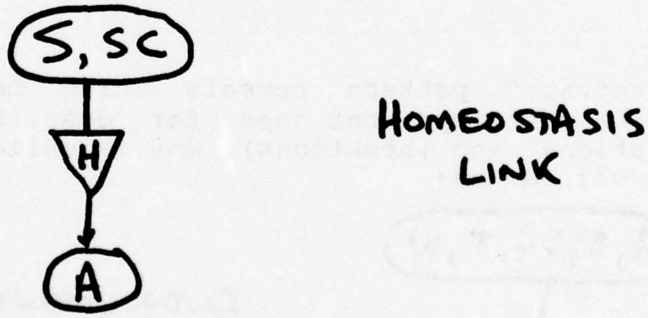
This link allows CSA representation to express the relationship of an external event to internal mental or emotional states which that event might "induce" within a perceiver of the event. Thus, for example, to capture the principle "If P loves Q and P sees R kiss Q, P might experience an induced state of feeling jealousy toward R" we write:**



Pattern 5

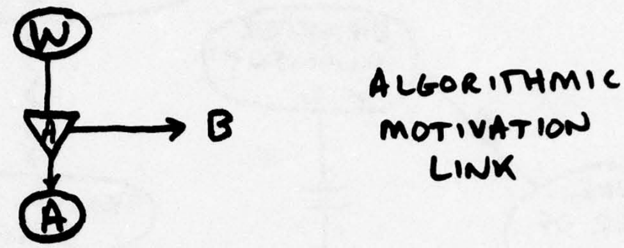
** Many of the event predicates we use in CSA have been adapted from Schank's conceptual dependency theory [Schank, 1972]; however, CSA makes no assumptions concerning whether or not such predicates are psychological primitives.

The second new CSA link employed in the contract pattern is written:



and is called the HOMEOSTASIS link. This link ties internal and mental states of a potential actor to predictions about actions he might perform to compensate for such states. The homeostasis link models an inherently non-algorithmic process (i.e. why does some internal state of a potential actor motivate that actor to perform an action?). After all, at some point, we must "cut" the CSA model of cause and effect and say simply "because it's the way a human is defined". However, although the homeostasis link's basis is inherently non-algorithmic, its role in a CSA pattern can be highly algorithmic: if P needs to arouse in Q a feeling of jealousy, some pattern containing a homeostasis link might tell P that one potentially fruitful tactic is to achieve some external event in the world, making sure that Q is aware of it! The point, of course, is that even though the basis of cause and effect in the psychological domain is hard to identify, it can still be described and put to use in algorithmic ways.

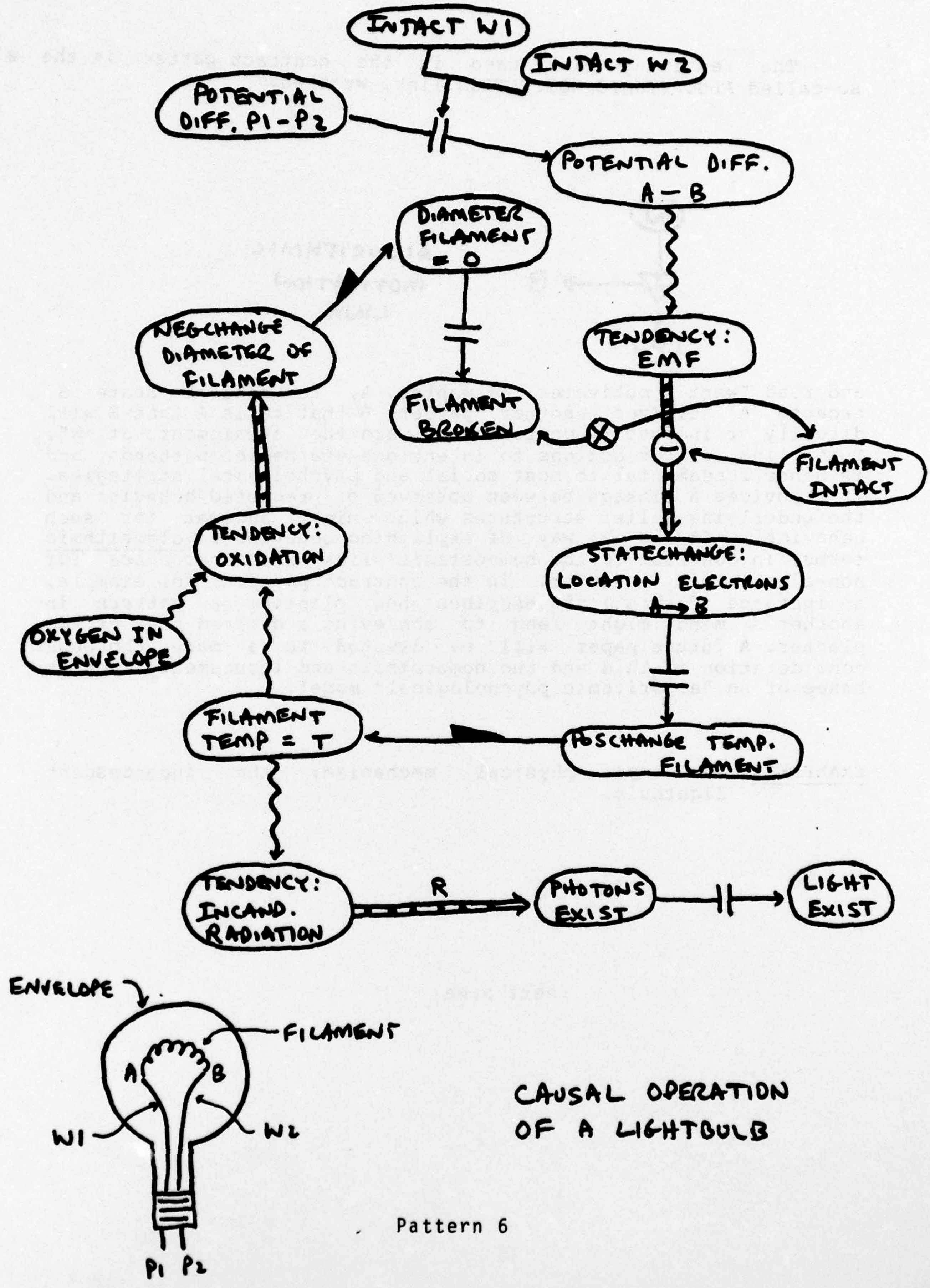
The remaining link used in the contract pattern is the so-called ALGORITHMIC MOTIVATION link, written:



and read "want W motivates the wanter, A, to achieve state S, because A believes another pattern B that tells A that S will directly or indirectly contribute toward the attainment of W". This link relates actions to intentions via belief patterns, and is hence fundamental to most social and psychological strategies. It provides a linkage between observed or predicted behavior and the underlying belief structures which might account for such behavior. It is a way of explaining behavior in <u>algorithmic</u> terms, in contrast to the homeostasis link which accounts for non-algorithmic behavior. In the contract pattern, for example, an instance of this link describes how planting a pattern in another's mind might lead to achieving a desired goal of the planter. A future paper will be devoted to a more thorough consideration of this and the homeostasis and inducement links as bases of an "algorithmic psychological" model.

<u>EXAMPLE 3</u>: A simple physical mechanism: the incandescent lightbulb.

(next page)


INTACT W1
INTACT W2
POTENTIAL DIFF. P1 - P2
POTENTIAL DIFF. A - B
DIAMETER FILAMENT = 0
NEGCHANGE DIAMETER OF FILAMENT
TENDENCY: EMF
FILAMENT BROKEN
FILAMENT INTACT
TENDENCY: OXIDATION
STATECHANGE: LOCATION ELECTRONS A → B
OXYGEN IN ENVELOPE
FILAMENT TEMP = T
POSCHANGE TEMP. FILAMENT
TENDENCY: INCAND. RADIATION
R
PHOTONS EXIST
LIGHT EXIST
ENVELOPE
FILAMENT
A
B
W1
W2
P1 P2
CAUSAL OPERATION OF A LIGHTBULB

Pattern 6

This pattern would be stored under "how to make light" in the larger organization of the memory. It illustrates how CSA represents a "causal topology" of physical devices, in contrast to their physical topology. We have so far employed the CSA representation to describe articles from the pencil (rather simple physically, rather complex in its causal structure) to the computer (complex in both domains), and it is in this domain of physical devices that CSA seems most complete as it is now defined.

This pattern about incandescent lightbulbs is read roughly as follows:

> "A potential difference across P1 and P2 (creating this potential difference relates to another mechanism description about switches, stored somewhere else in the memory) will be synonymous with a potential difference across points A and B (referring to the diagram), providing that wires W1 and W2 are intact. A potential difference across A and B continuously enables the tendency EMF (voltage), an actor-like entity. EMF will cause electrons to move from A to B through the filament, providing that the filament is intact. This current through the filament will be synonymous with an increase in filament temperature. (Note that the representation allows us to omit a description of how this occurs if we choose not to describe the principle of resistance in detail). This increase in filament temperature will eventually result in the filament temperature reaching some threshold T, which will then provide continuous enablement to two other tendencies: OXIDATION and INCANDESCENT RADIATION. INCANDESCENT RADIATION will repeatedly cause photons to exist, which is simply another way of saying that light exists, the primary goal of the lightbulb's operation. Also, provided there is oxygen present in the envelope (another continuous enablement required by OXIDATION), OXIDATION, viewed as another actor, will cause a continuous decrease in the diameter of the filament (eating it away). Eventually, the filament's diamenter will become zero, which is simply another way of saying that the filament is no longer intact. When this happens, the causality from EMF that is moving the electrons through the filament will be severed, and the lightbulb will shut down (i.e. burn out)!"

The new link present in this pattern is the ANTAGONISTIC STATE MARKER, written as

ANTAGONISTIC STATE MARKER

provides a way to highlight feedback loops in CSA patterns. The interpretation of this link is that the existence of S1 precludes the existence of state S2, and vice-versa. In other words, S1 and S2 are descriptions of mutually exclusive conditions. In a sense, this link is the inverse of the state coupling link. Its role in this mechanism is to relate (DIAMETER FILAMENT Ø) to (FILAMENT INTACT) as antagonistic states.

Once more we ask: what is the utility of a pattern like this which describes a physical or electrical device? Although this pattern is indexed (as are all other CSA patterns) from numerous places in the larger model in which it is stored, one of the primary indexings indicates that operating the incandescent light bulb is one way to cause light to exist. Again, such knowledge would be useful to either a plan synthesizer who itself required light, or to a story comprehender who was trying to understand a segment of a story where a knowledge of a lightbulb was central.

But clearly, we wouldn't want this pattern about a lightbulb to make a nuisance of itself if we were camping in the wilderness and needed light inside a cave! This rather whimsical observation leads us to the second part of the paper: selection of strategies and things on the basis of relevance within a given context or environment.

III. Selection

Imagining that we were to take a "snapshot" of the state of an individual's knowledge at some point, suppose that what we saw were thousands of CSA-like patterns containing the sorts of knowledge about cause and effect as have been suggested within the CSA representation. Suppose also that we were to see a large number of non-algorithmic entities representing concepts and tokens of concepts which modeled real-world entities such as ANIMAL, GREEN, JOHN SMITH, and so forth, as well as thousands of word concepts and their associated word sense concepts.

If we were then to observe any given act of, say, plan synthesis or language interpretation within the framework evident in this snapshot, we would notice a very remarkable thing: although there are probably millions of pieces of knowledge, the

human, viewed as either a plan synthesizer, a parser, or an interpreter of algorithmic activity about him, will seem to employ only a startlingly small fraction of this knowledge in accomplishing his task. Somehow, a knowledge of relevance seems to provide a tremendous filtering service to prevent a flood of irrelevant mental activity.

The phenomenon is pervasive. The same potent selectional force seems to be active wherever there is an element of choice involved: strategy selection, referent selection, "thing" selection, syntactic parse rule selection and word sense selection, to name only a few of the most obvious ones.

CSA theory maintains that this aspect of human intelligence - the apparent ability to filter out all but the most relevant knowledge at every point - is the most vital ingredient of so-called "intelligent" behavior, and that, however it happens, all forms of it are underlied by one cognitive mechanism. In the remainder of the paper, I want to consider what this mechanism of selection might look like, and how it might be roughly approximated by a computer.

Let us return first to consider the "observable" effects of selection. Strategy selection manifests itself as an agent which masks most of the overwhelmingly large number of alternative strategies for solving a given problem, making the problem solver "see" the most appropriate one first, or close to first. For example, suppose that P's goal is to go from the kitchen to the living room. Clearly, P will never even consider using a jet plane! Yet, as a pattern for getting places, except for certain relevance conditions, there is no a priori basis for avoiding this pattern. How, therefore, does selection rule out this pattern before the higher levels of the problem solver ever gain access to it?

There is the hint that, stored with every piece of knowledge, there is some sort of "user's manual" describing the conditions under which the piece of knowledge might be relevant. For strategies involving jet planes, the user's manual would indicate that such strategies are normally most applicable when distances are large, the plan executer has enough money, and so forth. Apparently it is this knowledge about knowledge with which the selection mechanism must deal, rather than with the knowledge itself. Let us call the knowledge itself "first-order knowledge", and the knowledge about knowledge "second-order knowledge".

So before P ever gets to consider the CSA pattern that tells him how to employ a jet to get him somewhere, some second-order knowledge about when this first-order knowledge is relevant apparently informs the selection mechanism not even to present the higher levels of the system with the jet strategies as alternatives. Furthermore, it is this second-order knowledge which must be most sensitive to context, since judgements of relevance are directly a function of the environment in which any

activity occurs.

Now, I want to make a point, but must take care not to get carried away. If overstated, the point might read like this:

> "Any system which makes intelligent selections at every decision point cannot be too far from being an accurate model of a human."

Of course, even if this were true, it would hinge entirely on what "good decision" means! A slightly more responsible conjecture is:

> "Any system that does not make intelligent selection at every decision point cannot be a good model of human intelligence."

A reformulation of this idea would be: It is more often a knowledge *about knowledge* that makes a system appear intelligent than it is the knowledge itself. Thus, even if our robot has crummy strategies for doing things, if it usually selects the best one for each task it attempts, I would still be willing to believe that it is behaving intelligently. Any system can know some particular strategy for moving an object. But the system will not appear "intelligent" unless it also knows when to apply this strategy in preference to all other possible strategies. By this standard, the measure of intelligence is: how well is the system able to select the most relevant strategy for each given task in a given context. In other words, "intelligence" is more a function of second-order knowledge than of first-order knowledge.

Strategy selection is a rather obvious form of selection. What other less obvious form are there? I want to point out three others, because, taken together with the strategy selection, these four forms of selection seem necessary to all forms of human symbolic intelligence.

The other three are: (2) event interpretation selection, (3) word sense selection during language comprehension, and (4) word choice during language generation.

We can define event interpretation to be that process which discovers how each perception relates to the context in which it is perceived. For example, how should we interpret the sentence: "John shouted at Mary"? Clearly, we ought to "see" different interpretations in different contexts: "John was on the opposite hilltop from Mary. John shouted at Mary." vs. "John was furious that Mary had stayed out so late. John shouted at Mary." and so on.

The third form of selection, word sense selection, is a well known problem in language analysis: it is the process of identifying an internal concept with a word of the language

spoken in context. Surprisingly, few computer-based parsers have dealt with problems of word sense selection; they eitner defer the problem by focussing on more syntactic issues, or simply ignore it because the domain of discourse for which the parser is designed permits them to, being narrow in scope... a "microworld" in the parlance of AI.

Representing the former camp, [Marcus, 1974] argues that identification of word senses can be bypassed in the initial phases of parsing, since, regardless of the sense, the syntax (e.g. case framework) of any given word is relatively fixed. If the syntactic component mislabels some case because it has ignored the word sense, so be it... it is nothing more serious than a mislabeling, because a subsequent semantic process will always know where to salvage the mislabeled case from the syntactic frame.

Although this point of view bothers me, I have yet to find a counter example to refute it; in fact, the CSA language front end interface behaves in this manner, using a version of Marcus' parser: CSA accepts syntactic case frames with Fillmore-like cases (as in [Fillmore, 1968]), then filters the frames through so-called "semantic discrimination networks" in order to map the syntax onto the meaning. If for example the sentences are: "John gave Mary a mean look.", "John gave Mary a teacup." and "John gave Mary an idea.", by testing the semantic types of the entities assigned to the various syntactic cases by Marcus' parser, the system will map these three thoughts - which look identical in syntactic structure - onto three quite different meaning structures:

(1) (C-INHEAD JOHN MARY (EFEEL JOHN ANGER MARY)) (John caused Mary to know that he felt anger toward her)

(2) (CSC-POSSESSION JOHN TEACUP JOHN MARY) (John caused a statechange in physical possession of the teacup from himself to Mary)

(3) (C-INHEAD JOHN MARY IDEA) (John caused some idea to be in Mary's mind)

The semantic discrimination networks which interface Marcus' parser to the CSA model also have access to expectancies in the system that are more than semantic, so that a sentence like "Mary picked the apple." will map onto "Mary indicated that it was the apple which she wanted." in one context, but "Mary plucked the apple." in another. A future paper on the CSA language component will describe how this occurs in more detail.

But I would argue that the artificial distinction between syntax and semantics (word sense selection) is not a good one. I

personally feel that, although Marcus' parser is perhaps the best-conceived parser around, syntax should not be done all at once as a preprocess, with the result being handed in a lump to semantics, as it is in our current system. Rather, I think a more accurate model of human parsing would be one that starts with semantics, having semantics (e.g. those questions posed by the semantic nets in the existing CSA language interface) call the syntactic component to answer semantically motivated questions such as: "is the semantic category of the sentence's direct object 'human', 'location' or 'mental-concept'?". This would amount to "syntax on demand", something a little less radical than the approach to parsing advocated by [Riesbeck, 1974]. The point of this approach is that, while the existence of a syntactic component is still acknowledged, only as much syntax as required by the meaning extraction process would be dealt with, rather than attempting to construct a complete syntactic analysis before any interaction with semantics. What this has to do with intelligent selection will become clearer in a few moments.

The fourth form of selection mentioned earlier, word choice during generation of language, will turn out to be an approximate inverse of the word sense selection process, relying on the same knowledge about knowledge that the sense selection process requires.

With this motivation, we ought now to ask: how is knowledge about knowledge to be stored? The requirements of such knowledge are now a little clearer: it must be capable of fueling higher level processes (strategy selection, word sense selection, perception interpretation and word choice during generation) with only the most relevant options, masking all else. During strategy selection, this will cause only the most relevant approach to a problem to be seen; during parsing, this will cause only the most relevant word sense for each word in context to be seen by the parser as it extracts thoughts from the language.

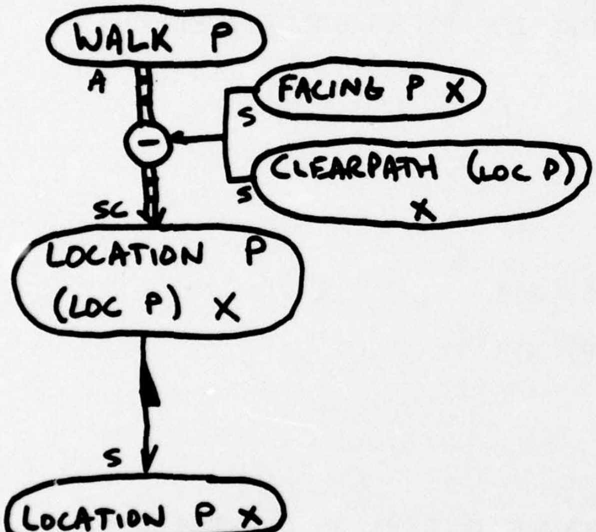
Let us now return to the user's manual metaphor. Suppose that every piece of knowledge has a user's manual. The user's manual for strategies will tell the plan synthesizer which strategy is likely to be most relevant for solving any given goal in a particular context. The user's manual for each word sense will tell the semantics, which are trying to map a syntactic case framework onto a CSA meaning structure, which senses of words to select. How is it that all this "user information" is coordinated? How is it organized?

The CSA theory proposes that, as each piece of knowledge enters the system, it is dissected into two pieces: the first order knowledge and the second order user's manual. The user's manual is taken apart and integrated into a so-called selection network.

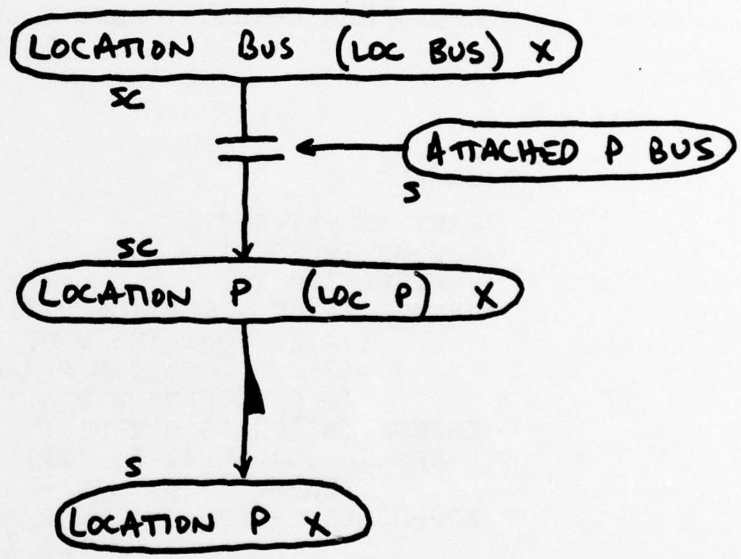
A selection network is an n-way branching discrimination

network, consisting of a connected collection of nodes. Each node contains a test and a set of alterative branches to be followed on the basis of the test outcome. Tests in the CSA system are query templates which are presented to the CSA database and deductive components (described in [Rieger, 1976b]) as the selection network is "applied". Applying a selection network means to consult it, threading a path through pieces of user's manuals which have been implanted in an organized fashion in the network nodes, until a result is reached. A result is, depending on the type of the network, a strategy, a word-sense, or an entity of whatever type it is that is being selected. The purpose of selection networks is to serve as a central unifying structure which will serve as an "intelligent arbiter".

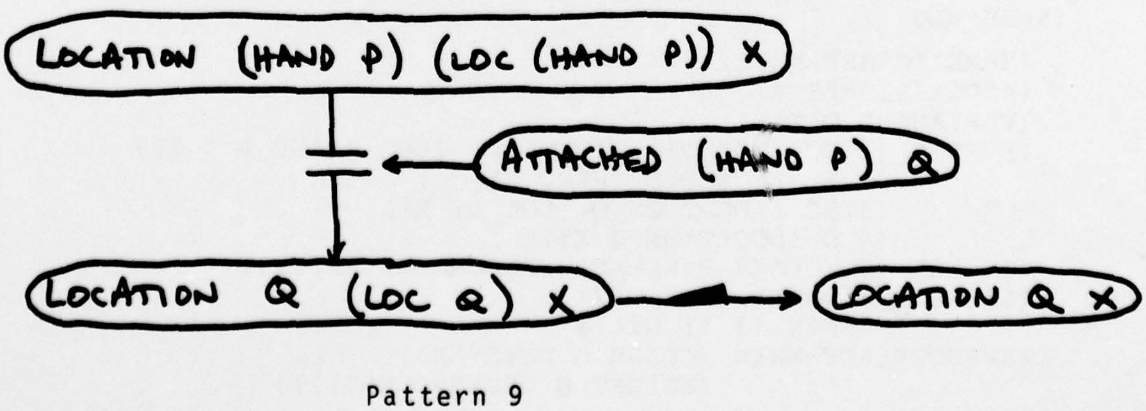
Let us now look at two very simple case studies in selection networks. In the first one, we will give the CSA system three strategies for moving an object: two dealing with humans, and one dealing with small, graspable objects. The three patterns will roughly approximate the notions of "walk", "take a bus" and "pick it up with your hand":



Pattern 7



Pattern 8



Pattern 9

When talking to the CSA computer model, we communicate these patterns to the system as follows:

Pattern 7:

```
($ABS-ALG (
  (NAME *WALK)
  (VARIABLES P X)
  (ACCOMPLISHES 5)
  (EVENTS (1 A (WALK P))
          (2 S (FACING P X))
          (3 S (CLEARPATH (LOC P) X))
          (4 SC (LOCATION P (LOC P) X))
          (5 S (LOCATION P X)))
  (THINGS)
  (LINKS (C-CAUSE (1 4) (2 3))
         (THRESH (4 5) NIL))
  (APPROPRIATE-WHEN (CLASS P HUMAN)
                    (LESS (DISTANCE (LOC P) X) ORDERMILE))))
```

Pattern 8:

```
($ABS-ALG (
  (NAME *GO-BY-BUS)
  (ACCOMPLISHES 4)
  (VARIABLES P X)
  (EVENTS (1 SC (LOCATION B (LOC B) X))
          (2 S (CONTAINS B P))
          (3 SC (LOCATION P (LOC P) X))
          (4 S (LOCATION P X)))
  (THINGS (B (CLASS B BUS)))
  (LINKS (S-COUPLE (1 3) (2))
         (THRESH (3 4) NIL))
  (APPROPRIATE-WHEN (CLASS P HUMAN)
                    (GREATER (DISTANCE (LOC P) X) ORDERMILE))))
```

Pattern 9:

```
($ABS-ALG (
  (NAME *GRASP-MOVE)
  (ACCOMPLISHES 4)
  (VARIABLES Q X)
  (EVENTS (1 SC (LOCATION (HAND P) (LOC (HAND P)) X))
          (2 S (ATTACHED (HAND P) Q))
          (3 SC (LOCATION Q (LOC Q) X))
          (4 S (LOCATION Q X)))
  (THINGS (P (CLASS P HUMAN) (RECOMMEND SELF)))
  (LINKS (S-COUPLE (1 3) (2))
         (THRESH (3 4) NIL))
  (APPROPRIATE-WHEN (CLASS Q PHYS-OBJ)
                    (WEIGHT Q ORDERPOUNDS))))
```

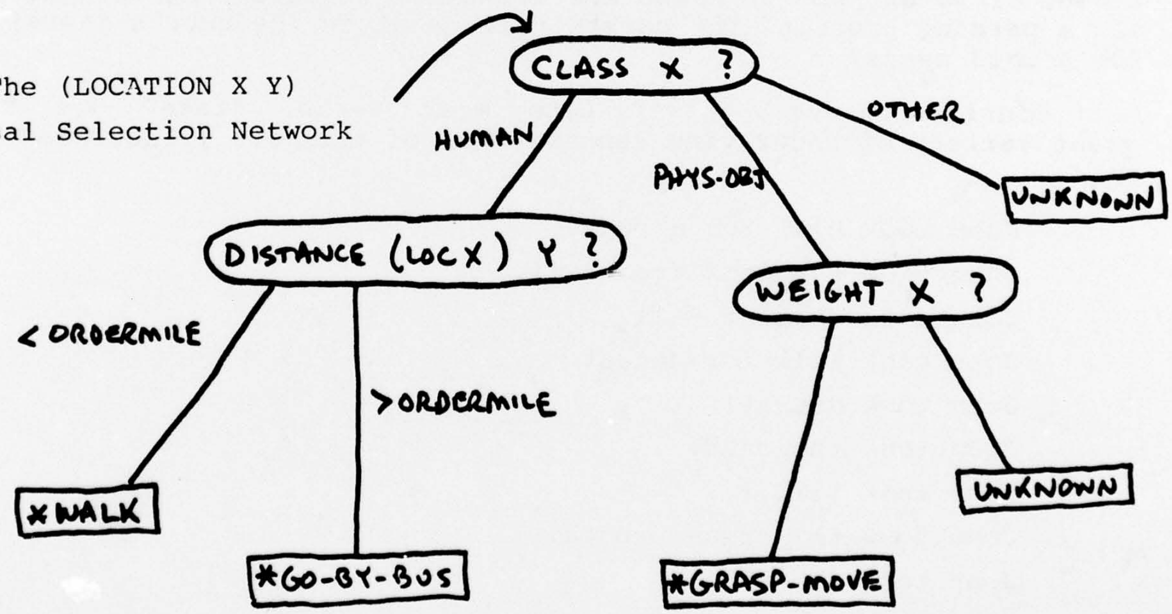
Selection networks for strategies are cataloged by the primary predicate describing the state of the world the strategy is intended to achieve. In these cases, the predicate is LOCATION. Hence, the user's manuals for all three of these patterns will be synthesized into the (initially empty) AGENT W CAUSES STATECHANGE LOCATION X Y Z strategy selection network. (In CSA, we call these networks "causal selection networks".) In a large CSA system, there will be a rather complex causal selection network for each state and statechange predicate known to the problem solver.

In the CSA computer model's syntax, the user's manual is signaled by the keyword APPROPRIATE-WHEN. The information associated with this keyword is in the form of statements about conditions which the variables in the strategy must satisfy, or statements about conditions which must be true (i.e. in the CSA database) at the time the strategy is being selected. The "appropriate-when" conditions are taken apart and used to construct an initial selection network. How this occurs is a matter of considerable theoretical interest, since we believe it represents a significant form of learning. However, these issues will not be discussed here.

In this example, the network which results from the synthesis of these three user's manuals will look like:

The (LOCATION X Y)

Causal Selection Network



Now, whenever the system is confronted with a goal of the

form: "construct a plan wherein agent W causes a statechange in X's location from Y to Z", this causal selection network will be called up and "applied". The data base and deductive components of the system will subsequently see a progression of queries about various features of the W, X, Y and Z, and about the general state of the world, until the network finally chooses one of the three strategies as most relevant, or determines that it does not have a relevant strategy for the given goal, according to those criteria.

Now we have an arbiter, which has been built up automatically from the user's manuals of the various strategies among which it will select. This arbiter will be the agent which performs the crucial pre-filtering of strategies for the higher levels of the system. Once the plan synthesizer commits itself to a particular filtered strategy, the strategy will communicate a set of subgoals to the synthesizer, and these subgoals will evoke recursive behavior for each subgoal identical to the top level behavior. (Actually, there are some other processes which enter the picture as subgoal solutions are constructed. Among such processes are "demons" which will protect a subgoal once it has been solved.)

As a second case study in selection networks, let us consider an example of <u>thing</u> selection in which the things are senses of words, and in which the selection is occurring as part of a parsing process. The question is: what is the user's manual for a word sense?

Consider the verb "take". Like most verbs, "take" has a great variety of underlying senses. Some of them are illustrated by:

John took Bill for a ride.

John took the book from Bill.

John took care of Bill.

John took Bill for honest.

John took drugs.

John took the oath.

John took Bill.

John took the green banana.

John took a break.

John took for the hills.

John took up the guitar.

These senses are the counterparts of the strategies in strategy selection. The user's manual for each sense consists of a set of constraints at all the various levels of language: the <u>lexical</u> and <u>grammatical</u> context in which the sense may be used, a <u>set of</u>

semantic constraints on the types of case-fillers the sense accepts, and, most important (since it ties the parse process in with general world knowdeledge and "deep" comprehension processes), contextual constraints on the types of situations in which the word sense might be used.

For "take", an example of lexical environment is: one of the senses of "take" meaning "to begin a habitual activity" or "to reel in" or "to agree to" (among possibly others) is suggested when the lexical item immediately to the right of take is "up". An example of a grammatical rule is: if there is a prepositional phrase beginning with "to" and specifying a location, then "take" might have the interpretation "to move toward rapidly", as in "take to the hills". An example of contextual environment is: if the actor associated with the verb "take" is expected to exhibit selection behavior (e.g. to select which apple to eat) then the sense of "take" meaning "to select" is particularly appropriate (as in "Mary took the green apple.").

By applying word sense networks from the bottom up, it should also be possible to generate language. That is, by starting from an internal concept (word sense) that requires expression, and climbing the sense selection network in which that concept appears as a terminal node, the generator would do whatever was required to *cause* the *answers* to the network selection questions be true. Doing so would spell out all relevant aspects of the linguistic and conceptual environment surrounding the word thus selected.

A partial word sense selection network for the verb "take" might look as follows:

(next page)

questions posed by networks when they are applied, i.e., when parsing is performed.

A parser which reflected this theory would therefore be little more than a central control for the application of word sense networks, one such network representing each word in the sentence being parsed. All would be run essentially in parallel, each attempting to discover the most likely sense of the word it represents in the sentence. We have not yet developed this notion as a computational model, but would expect a key issue to be how the various networks cross- communicate during the parse, transferring information among one another. But regardless of how such a parser would actually function, syntax would be performed only as far as the questions posed by the sense selection networks demand to perform their job.

This theory raises many intellectual issues which we will not address here. The main one is: is it really reasonable to "multiply out the grammar" by distributing knowledge about language across the individual words? Is it not more reasonable to concentrate on intelligent selection of the factored form of language, i.e. the grammar, which attempts to express the general principles of the language's structure? While this is the traditional point of view, I personally see far more potential for learning and adaptive behavior in the word sense network approach. The grammar can come later. We often fail to realize that words really are individuals, with very specific, often complex world knowledge associated with them; they are not simply members of an abstract syntactic category referenced by some grammar. Why put more emphasis on grammar than anything else? I submit that making grammar central not only is an artificial way to slice through language, but it is also an incorrect one that leads one into incredibly baroque theories of abstract grammar which have no practical value, in the sense that computational models could be constructed from them. Perhaps by turning the traditional approach to language inside out (which is how I imagine the word sense selection system), things won't be so difficult!

IV: Conclusion

Perhaps it is time to pop back up to the surface and conclude. The points of this paper have been twofold: first, that it is important when dealing with language to have a well-defined and concise theory of how to represent general world knowledge, and second that being able to make intelligent selections among alternatives in this knowledge is at least as important as the knowledge itself. Since these two issues pervade all aspects of language understanding and problem solving, they will remain with us for quite a while.

References

[Abelson, 1973] Abelson, R., "The Structure of Belief Systems," in Schank & Colby (eds.), Computer Models of Thought and Language, W.H. Freeman, 1973

[Fillmore, 1968] Fillmore, C., "The Case for Case," in Bach & Harms (eds.), Universals in Linguistic Theory, Holt, Rinehart & Winston, 1968

[Marcus, 1974] Marcus, M., "Wait and See Strategies for Parsing Natural Language," M.I.T. A.I. Lab Working Paper 75, 1974

[Rieger, 1975] Rieger, C., "The Commonsense Algorithm as a Basis of Computer Models of Human Memory, Inference, Belief and Contextual Language Comprehension, in Proc. Workshop on Theoretical Issues in Language Processing, M.I.T., 1975

[Rieger, 1976a] Rieger, C., "A Representation of Knowledge for Problem Solving and Language Comprehension," Artificial Intelligence, vol. 7, no. 2, Summer 1976

[Rieger, 1976b] Rieger, C., "Spontaneous Computation in Cognitive Models," Univ. of Maryland TR 459, 1976

[Riesbeck, 1974] Riesbeck, C., "Computational Understanding: Analysis of Sentences and Context," Ph.D. dissertation, Stanford Univ. 1974, AI Memo 238

[Schank, 1972] Schank, R., "Conceptual Dependency: A Theory of Natural Language Understanding," Cognitive Psychology, 3(4), 1972

[Disney, 1975] ______, The Magic Grinder, Walt Disney Productions, Random House, 1975